Efficacy analysis of Zearn Math: Findings from implementation in Louisiana

May 2023

Jessica Rickel, M.A.
Alisa Szatrowski, Ph.D.
Charlie Rosemond, M.S.
Zearn

Abstract

Using Coarsened Exact Matching (CEM), this analysis examined the impact of Zearn Math across 7,116 matched pairs of students in 31 Louisiana parishes. Students who completed an average of 3+ Zearn Math lessons per week during the 2021-2022 school year were compared to similarly matched peers who completed less than 1 lesson per week.

Findings showed that, on average, students who used Zearn Math grew an additional 6.3 scale score points on the Louisiana Educational Assessment Program (LEAP) compared to non-users. Students who started the year below proficiency gained an additional 8.2 scale score points, the equivalent of 1.3 years of learning, compared to non-users. Consistent Zearn Math users were also more likely to improve their Achievement Level, had a higher percentage of students placing at the “Mastery” level or above in spring 2022, and had higher probabilities of students meeting their 2022 target score.

Table of Contents

List of Tables

Introduction

Zearn is the 501(c)(3) nonprofit educational organization behind Zearn Math, a top-rated math learning platform used by one in four elementary-school students and by more than 1 million middle-school students nationwide. This report summarizes findings from an efficacy analysis of Zearn Math, which is implemented in select parishes in the state of Louisiana. The goal of this study was to isolate the impact of Zearn Math on student achievement through quasi-experimental matching methods that facilitate causal inference.

This efficacy analysis was conducted across 31 parishes in the state of Louisiana. Across these 31 parishes, there are 370,555 total students, of whom 108,508 are in Grades 4 to 7. The student body is 52% Black and/or Latino, 71% economically disadvantaged, 3% multilingual learners (MLLs), 13% students in special education, less than 5% gifted or talented, and 19% chronically absent (Louisiana Department of Education, 2022a, 2022b, 2022c).

In Grades 4 to 7, there were 7,679 consistent Zearn Math users — those who completed 3+ lessons per week or approximately 90+ lessons per year — who could be matched to assessment data from the 2020-2021 and 2021-2022 school year (see Appendix A Tables A1 and A2 for a breakdown of sample demographics and comparison between sample and statewide demographics).[1]

This study was designed to meet the What Works Clearinghouse (WWC) “Meets WWC Group Design Standards with Reservations” rating and to meet an Every Student Succeeds Act (ESSA) Tier 2 (Moderate) rating on the ESSA guidelines for evidence-based interventions. The study uses quasi-experimental matching methods to create baseline equivalency between treatment and control groups along major confounding factors (see Appendix B for more information).

Methodology

Quasi-experimental matching techniques were used to isolate the impact of Zearn Math on student achievement across Louisiana. Consistent Zearn Math users were matched to non-users[2] on starting math and English Language Arts (ELA) Louisiana Educational Assessment Program (LEAP) scores, along with ten student characteristics. The goal of matching was to create 1:1 pairings between similar students, differing primarily on Zearn Math usage during the 2021-2022 school year. The outcome under investigation was the average treatment effect on the treated as controls were selected to match

[1] The sample population of fidelity users differs from the total population of these 31 parishes, having a proportionally smaller population of students from disadvantaged backgrounds. The implications of this difference are discussed in the limitations section.

[2] Non-users are those who completed less than one lesson per week.

individuals in the treatment group.

In order to see maximum benefit from Zearn Math, students are advised to complete three or more digital lessons per week during the school year. Therefore, the treatment group was composed of students who consistently used Zearn Math during the 2021-2022 school year, operationalized as an average of three or more digital lessons per week, i.e., 90 or more digital lessons per year. The control group was selected from other students in Louisiana with no Zearn Math usage, operationalized as an average of fewer than one digital lesson per week, i.e., fewer than 30 digital lessons per year.[3]

Drawing causal inference from observational data is challenging because factors that impact a person's likelihood to receive an intervention may also impact their outcomes. Therefore the differences in outcomes observed between individuals may not be caused by the intervention itself, but by other confounding factors that imbalance the treatment and control groups (Stuart, 2008; Iacus et al., 2011).

Matching methods were used to balance the composition of confounding factors between individuals who consistently used Zearn Math (the treatment group) and a comparison group of individuals who had no Zearn Math usage (the control group). This is done to isolate the difference in outcomes from the intervention itself, separate from any impact due to potential confounding factors.

This efficacy analysis used a two-step Coarsened Exact Matching (CEM) method with optimal matching to create a control group that was as similar as possible to the treatment group of consistent Zearn Math users. CEM is a technique that simulates block sampling by matching students on covariates, demographic and academic factors that may be related both to a student's likelihood of using Zearn Math consistently and their academic performance (Blackwell et al., 2009; Iacus et al., 2011). The effectiveness of matching is conditional on the ability of observable factors to capture the selection process that sorted individuals into treatment and control. Models that do not capture major factors may produce biased estimates.[4]

Using CEM, treatment students were put into matching strata with control students that were in the same grade and within 5 scale score points on both the math and ELA spring 2021 LEAP. Then, within

[3] This definition of treatment and control does not use an intention-to-treat (ITT) framework that would include in the treatment all students that had been offered Zearn Math (McCoy, 2017). While the ITT approach is the most efficacious for identifying the impact of a program under real-world implementation constraints, the goal for this study was to understand the impact of fidelity usage in the hopes of increasing fidelity usage of the platform across schools. This efficacy analysis examines the impact of Zearn Math, implemented with fidelity, vs. with low to no usage. The implications of Zearn's approach are discussed further in the limitations section.

[4] This potential for bias does not exclude a study from meeting WWC's "Meets WWC Group Design Standards with Reservations" rating as long as baseline equivalency can be established. According to WWC: "In QED studies, confounding is almost always a potential issue due to the selection of a sample, because some unobserved factors may have contributed to the outcome. The WWC accounts for this issue by not allowing a QED study to receive the highest rating" (What Works Clearinghouse, 2020).

strata, treatment students were matched to control students with whom they shared at least six of ten other demographic and academic characteristics: parish, school, use case,[5] gender, race, economic disadvantage,[6] multilingual learner (MLL) status, special education status, gifted status, and chronic absenteeism status.[7]

This optimal matching method utilized Bertsekas' auction algorithm to produce combinatorial optimization such that treatment individuals were matched to others closest to them in the control pool. When controls were the best-fit match for more than one treatment individual, the pairing went to the individual from whom the next best pairing was the farthest (Bertsekas, 1981; Rosenbaum, 2020).[8]

If a treatment student had no match within their grade and score strata with whom they shared at least six characteristics, they were excluded from the analysis. The caliper that limited match difference to no more than four characteristics was selected to maximize inclusion in the sample, prevent biasing through uneven patterns of exclusion, and still ensure similarity between groups.

For more information on Zearn's methodological approach, see ***Efficacy Analysis Methodology: Zearn's approach to Coarsened Exact Matching***.

Difference of Means

Once consistent Zearn Math users were matched to a similar group of non-users, a difference of means analysis was conducted to quantify the impact of Zearn Math on student achievement. Means were calculated for treatment and control groups overall as well as for groups disaggregated by starting math Achievement Level and demographic factors. Because pairs of consistent Zearn Math users and non-users were allowed to mismatch on up to four demographic characteristics, subgroups did not always align on starting scale scores. Therefore, differences in achievement by demographic subgroup

[5] Use case refers to whether the parish a student is in uses Zearn Math as its core curriculum or as a digital complement.

[6] Student-level economic disadvantage data was not available. Instead, the percent of economically disadvantaged students in a grade level within a school was used to classify students into low (0-40%), mid (40-75%), and high (75%+) poverty groups. These thresholds were chosen to align with NCES (2022). The economically disadvantaged subgroup is composed of students who are in a school where 75% or more students in their grade are considered economically disadvantaged.

[7] Chronic absenteeism is defined by Louisiana as missing 10% or more school days.

[8] In other words, if Control Student A was the best match for Treatment Student 1 and Treatment Student 2, sharing 6 out of 7 characteristics with each, Control Student A could still only be matched with either Treatment Student 1 or Treatment Student 2. If the next best match for Treatment Student 1, Control Student B, shared 4 characteristics, and the next best match for Treatment Student 2, Control Student C, shared 5 characteristics, then Treatment Student 1 would be matched with Control A and Treatment Student 2 would be matched with Control C. In this way, the algorithm of optimal matching balances the closeness of any individual match with its impact on the closeness of the overall group match.

were reported as difference-in-difference, rather than as raw scores.

Academic growth was measured as the change in math scores between the spring 2021 and spring 2022 LEAP administrations. LEAP has five Achievement Levels: Unsatisfactory (1), Approaching Basic (2), Basic (3), Mastery (4), and Advanced (5). Students scoring at "Mastery" or above are considered proficient (Louisiana Department of Education, 2021). Outcomes are reported in terms of change in scale score, change in Achievement Level, percent placing at "Mastery" or above on the spring 2022 LEAP, and change in this percent between spring 2021 and spring 2022.

Difference in means *t*-tests were run on the average academic growth of the treatment group vs. the average academic growth of the control group to determine if the impact of Zearn Math was statistically significant. Given *SD*=standard deviations and *n*=number of observations per group, *t*-tests were conducted as:

$$t = \frac{mean_{treatment} - mean_{control}}{\sqrt{\frac{SD^2_{treatment}}{n_{treatment}} + \frac{SD^2_{control}}{n_{control}}}}$$

Effect size was calculated with *Cohen's d*, which divides the difference in means between treatment and control by the pooled standard deviations:

$$Cohen's\ d = \frac{mean_{treatment} - mean_{control}}{pooled\ SD}$$

Yearly Growth

In addition to growth in scale score points, growth among consistent Zearn Math users and non-users was translated into years of learning for those who placed below proficiency ("Unsatisfactory," "Approaching Basic," or "Basic" level) on the spring 2021 LEAP.

For all students who place below proficiency on the LEAP, Louisiana provides a "Growth to Mastery Target" that represents the number of scale score points a student is expected to increase each year in order to achieve "Mastery" by 8th grade. The "Growth to Mastery Target" is not calculated for students starting at the "Mastery" or "Advanced" levels, as this target intends to provide below-proficient students a pathway to proficiency by the 8th grade, and these students are already on track. Instead, the target scores for these students are based on the goal of achieving or maintaining "Advanced" placement.

Louisiana targets 8th-grade math proficiency for its multiyear benchmark as this represents on-time Algebra readiness, the mastery of skills necessary for students to pass Algebra I their first time in the course. Algebra readiness is seen as a key benchmark in the field, as passing Algebra on time gives

students access to higher-level math and science courses, and is shown to improve the likelihood of graduating high school, attending college, and graduating college (Evan et al., 2006; National Mathematics Advisory Panel, 2008).

For the purpose of this analysis, one year's growth was defined as the number of points a student was expected to gain each year, based on their "Growth to Mastery Target." Yearly growth was also calculated for each starting Achievement Level and differs from the total below-proficient sample, because students who start at the lowest Achievement Level are expected to grow more each year to be on track for 8th-grade proficiency than those at the "Basic" level.[9]

Because students scoring at the "Mastery" and "Advanced" levels are not provided traditional growth targets, but rather expected to maintain proficiency, students at these levels are not included in the calculation of yearly growth.[10]

Probability Models

Additionally, logistic regression models calculated the difference in probability of students meeting their math target score on the spring 2022 LEAP between consistent Zearn Math users and non-users.[11] Outcomes are also expressed as relative likelihoods, i.e., the probability of the outcome among consistent Zearn Math users vs. non-users across all students, as well as disaggregated by starting Achievement Level and subgroup.[12]

[9] Louisiana's "Growth to Mastery Target" is calculated by subtracting a student's scale score from 750 (the minimum score needed to reach "Mastery" in 8th grade) and then dividing that by the number of years away from 8th grade. For example, a third-grade student scoring a 700 would need to earn 50 points by 8th grade to score a 750. Since they are five years away, they are expected to earn 10 points per year to stay on track, meaning their 2022 target score is 710 (Louisiana Department of Education, 2018).

[10] The target score for "Mastery" students is calculated similarly to the "Growth to Mastery Target" (see footnote 9), except the 2021 scale score would be subtracted from 801 (the minimum score for "Advanced" in the 8th grade). However, if the 2021 score exceeds the "Advanced" score for the next grade or the points per year needed to reach "Advanced" in 8th grade are less than the points needed to reach it by the next grade, then the student's target score is the minimum "Advanced" score for the next grade. "Advanced" students' target is to maintain this placement (Louisiana Department of Education, 2018). It is possible that a student's 2022 target score is less than their 2021 score, so "expected growth" would be negative, meaning it cannot be accurately translated into years of learning.

[11] For students starting below proficiency, the target score is based on the "Growth to Mastery Target" calculation (see footnote 9). For those starting at or above proficiency, the target score is based on achieving or maintaining "Advanced" (see footnote 10).

[12] Because the groups have been matched on students' starting scale scores, grade, and ten academic and demographic factors, these variables are already controlled in the models.

CEM One-Year Impact Analysis

Out of Louisiana's 7,679 consistent Zearn Math users, all but 563 were matched, creating a sample of 7,116 matched pairs. Treatment and control groups differed by an average of 2.16 demographic factors, 2.61 in starting math, and 2.41 in starting ELA scale score points.[13] The 563 consistently using students excluded from the study due to lack of match did not concentrate in any demographic category that would bias the sample (see Appendix A Table A1 for a breakdown of sample demographics). Although there were slightly amplified differences in starting scores among some demographic subgroups, all groups met baseline equivalency to be included in this analysis.[14]

Results

Across all students and each starting Achievement Level, consistent Zearn Math users gained more scale score points than matched non-users. On average, consistent Zearn Math users gained 5.6 scale score points, while matched non-users lost 0.6 of a scale score point between the spring 2021 and spring 2022 LEAP administrations, a difference of 6.3 points (effect size = .27; see Appendix A Table A6 for findings from the difference in means analysis and Results Table 1).

RESULTS TABLE 1

Growth Across LEAP Achievement Levels

Scale score gains for consistent Zearn Math users (Treatment) vs. non-users (Control), by starting Achievement Level*

	Consistent Users	Non-Users	Difference
All Students	5.6	-0.6	6.3
Unsatisfactory (Level 1)	22.5	11.8	10.7
Approaching Basic (Level 2)	12.4	2.9	9.4
Basic (Level 3)	5.8	-1.0	6.8
Mastery and above (Levels 4 and 5)	1.2	-3.3	4.5

*Score growth is greater for students starting below proficiency because they have more room to grow. It is also possible that a "Mastery" score can exceed the lowest score for "Advanced" in the next grade level, meaning a student may demonstrate better performance and be placed at a higher Achievement Level in 2022, despite having a lower scale score than in 2021.

[13] Mean pretest math scores differed by 0.3 points. This is less than .05 of a standard deviation of the combined means. According to WWC, "baseline differences less than or equal to 0.05 standard deviations in absolute value automatically satisfy the baseline equivalence standard and do not require statistical adjustment" (WWC, 2022, p. 53). See Appendix A Table A3 for full details on baseline equivalence.

[14] Subgroups of female, male, and economically disadvantaged had baseline differences < .05 of a standard deviation, which satisfies baseline equivalence without adjustment, according to WWC. All other subgroups, with the exception of gifted status, had baseline differences < .25 of a standard deviation, satisfying baseline equivalency with a difference-in-difference adjustment (2022). The gifted status subgroup did not qualify for baseline equivalence with adjustment, therefore their results are not reported (see Appendix A Table A3 for full details on baseline equivalence).

For students who started below proficiency, scale score growth was translated into years of learning.[15] Students in the sample starting below proficiency were expected to grow 6.2 scale score points between the spring 2021 and spring 2022 LEAP administrations. Across all below-proficient students and for each starting Achievement Level, consistent Zearn Math users exceeded expectations, while non-users fell short. On average, consistent Zearn Math users gained 10.3 scale score points while non-users gained 2.1 scale score points, a difference of 8.2 points (effect size = .43; see Appendix A Table A6 for findings from the difference in means analysis). This translates to an additional 1.3 years of learning for consistent Zearn Math users relative to non-users, indicating Zearn Math users, on average, are on track to reaching proficiency by the 8th grade (see Results Table 2).[16]

RESULTS TABLE 2

Years of Growth for Below-Proficient Students

Scale score gains translated to years of growth for consistent Zearn Math users vs. non-users who started below proficiency, by starting Achievement Level*

	All Below-Proficient Students**	Unsatisfactory	Approaching Basic	Basic
Treatment growth in scale score	10.3	22.5	12.4	5.8
Treatment growth in years	1.7	1.6	1.5	2.0
Control growth in scale score	2.1	11.8	2.9	-1.0
Control growth in years	0.3	0.8	0.4	-0.3
Growth difference in scale score	8.2	10.7	9.4	6.8
Growth difference in years	1.3	0.7	1.2	2.4

*The average expected growth for all students starting below proficiency was 6.2 scale score points between spring 2021 and spring 2022. Disaggregated by starting Achievement Level, those who started at "Unsatisfactory" were expected to grow 14.3 scale score points per year, those who started at "Approaching Basic" were expected to grow 8.2 points per year, and those who started at "Basic" were expected to grow 2.9 points per year.
**This refers to the 3,482 treatment and 3,521 control students who started at the "Unsatisfactory," "Approaching Basic" or "Basic" Level.

Across all subgroups, consistent Zearn Math users gained scale score points between spring 2021 and spring 2022, while non-users lost points. Gains were even higher across traditionally disadvantaged subgroups of students including Black and/or Latino students, economically disadvantaged students,

[15] A description is available in the “Yearly Growth” section.

[16] Because LDOE calculates growth targets by taking a student's distance from “Mastery” and dividing it by the number of years they have until 8th grade, the lower a student starts academically, the more they are expected to grow each year to be on target. That is why, in Results Table 2, students who started at the “Unsatisfactory” Achievement Level grew more in points than those who started at the “Basic” Achievement Level, but had a lower years-of-growth equivalence.

MLLs, and students in special education. Notably, MLLs with consistent Zearn Math usage grew the most, gaining 7.9 scale score points, while matched non-users lost 0.7 of a scale score point, a difference of 8.6 points (effect size = .46; see Appendix A Table A7 for findings from the difference in means analysis and Results Table 3).

RESULTS TABLE 3

Growth Across Subgroups

Scale score gains for consistent Zearn Math users vs. non-users, by subgroup

	Consistent Users	Non-Users	Difference
All Students	5.6	-0.6	6.3
Female	4.7	-1.2	5.9
Male	6.5	-0.1	6.6
Black and/or Latino	6.7	-1.0	7.7
Economically disadvantaged	7.2	-1.5	8.7
MLL	7.9	-0.7	8.6
Special education	6.8	-1.4	8.3
Gifted	++	++	++
Chronically absent	5.5	-4.0	9.5

++Subgroup does not satisfy baseline equivalence even with statistical adjustment.

Mobility models compared the change in Achievement Level for treatment and control students based on their starting Achievement Level in spring 2021. Across all Achievement Levels, consistent Zearn Math users maintained or increased their Achievement Levels at higher rates than non-users. Notably, among consistent Zearn Math users who started at the “Unsatisfactory” level, 70% improved their Achievement Level, compared to only 45% of non-users. Similarly, among consistent Zearn Math users who started at the “Approaching Basic” level, 52% improved their Achievement Level, compared to only 33% of non-users (see Results Table 4).

RESULTS TABLE 4

Change in LEAP Achievement Level

Spring 2022 LEAP Achievement Level change relative to Spring 2021 LEAP Achievement Level				
Spring 2021 Achievement Level	**Zearn usage**	**Decrease**	**Maintain**	**Increase**
Unsatisfactory	Consistent Users	0%	30%	70%
	Non-Users	0%	55%	45%
Approaching Basic	Consistent Users	7%	41%	52%
	Non-Users	19%	48%	33%
Basic	Consistent Users	14%	50%	37%
	Non-Users	27%	48%	25%
Mastery and above*	Consistent Users	13%	87%	0%
	Non-Users	20%	80%	0%

Across all students and each subgroup, consistent Zearn Math users placed at “Mastery” or above at higher rates than non-users, despite having the same starting proficiency rates. Across all students, 55.8% of Zearn Math users placed at “Mastery” or above, compared to 48.4% of non-users, a difference of about 7.3%. Differences were even more pronounced for traditionally disadvantaged subgroups such as Black and/or Latino students, economically disadvantaged students, students in special education, and chronically absent students, with consistent Zearn Math users placing at “Mastery” or above at 1.2-1.4x the rate of non-users. The largest difference was for chronically absent students, of whom an additional 12% of consistent Zearn Math users placed at “Mastery” or above relative to non-users (see Results Table 5).[17] Further details are located in Appendix A Table A5.

[17] The percentages displayed in Results Table 5 reflect the mean-adjusted spring 2022 values. First, an adjusted spring 2021 baseline mean was calculated as the average percent of Zearn Math users and non-users starting at or above “Mastery.” Then the actual percentage-point change in those placing at or above “Mastery” between spring 2021 and 2022 was added to the adjusted baseline mean to obtain the adjusted spring 2022 percentages for treatment and control students. This corrected for subgroups that met baseline equivalence but did not have identical starting percentages of proficiency, so that outcomes could be compared accurately (see Appendix A Table A5 for the unadjusted and adjusted spring 2021 and 2022 values and further details).

RESULTS TABLE 5

Percent Reaching "Mastery" or Above Across Subgroups

Percent of students placing at "Mastery" or above on the spring 2022 LEAP for consistent Zearn Math users vs. non-users, by subgroup*

	Consistent Users	Non-Users	Difference	Relative Likelihood
All Students	55.75%	48.42%	7.34%	1.15
Female	52.69%	46.86%	5.83%	1.12
Male	58.71%	49.93%	8.78%	1.18
Black and/or Latino	38.00%	29.64%	8.37%	1.28
Economically disadvantaged	38.13%	30.70%	7.43%	1.24
MLL	**	**	**	**
Special education	25.89%	20.48%	5.41%	1.26
Gifted	++	++	++	++
Chronically absent	41.91%	30.24%	11.67%	1.39

*The values displayed reflect the adjusted spring 2022 percentages (see footnote 17 for a description of this calculation; see Appendix A Table A5 for further details).
**Excluded due to lack of statistical significance for change in the percent at "Mastery" or above. Full results are available in Appendix A Table A7.
++Subgroup does not satisfy baseline equivalence even with statistical adjustment.
NOTE: Summary report presentation of this data may vary slightly from detailed chart above due to rounding for visual simplification

Across all students and each subgroup, consistent Zearn Math users saw gains in the percent placing at "Mastery" or above on the LEAP between the spring 2021 and spring 2022, while non-users saw drops. On average, consistent Zearn Math users saw a 5.0 percentage-point increase in those placing at "Mastery" or above, while non-users saw a 2.4 percentage-point decrease, a difference of 7.3 percentage points. These differences were even larger across traditionally disadvantaged subgroups of students including Black and/or Latino students, economically disadvantaged students, and chronically absent students (see Results Table 6).[18] Full results are available in Appendix A Tables A6 and A7.

[18] For each subgroup in treatment and control, the percent placing at "Mastery" or above in spring 2021 was subtracted from the percent placing at "Mastery" or above in spring 2022. This change is depicted in Results Table 6. If the percent within a subgroup was the same in spring 2021 and spring 2022, the change listed in Results Table 6 would be 0. The spring 2021 and 2022 percentages are reported in Appendix A Table A5.

RESULTS TABLE 6

Change in Percent Placing at "Mastery" or Above Across Subgroups

Change in the percent of students placing at "Mastery" or above on the LEAP between spring 2021 and spring 2022 for consistent Zearn Math users vs. non-users, by subgroup

	Consistent Users	Non-Users
All Students	5.0%	-2.4%
Female	3.1%	-2.7%
Male	6.7%	-2.0%
Black and/or Latino	6.1%	-2.2%
Economically disadvantaged	4.6%	-2.8%
MLL	**	**
Special education	4.0%	-1.5%
Gifted	++	++
Chronically absent	5.4%	-6.3%

**Excluded due to lack of statistical significance. Full results are available in Appendix A Table A7.
++Subgroup does not satisfy baseline equivalence even with statistical adjustment.

Logistic Regression Results

Logistic regression models were calculated to obtain the differences in the probability of students meeting their spring 2022 LEAP target score. On average, consistent Zearn Math usage is associated with a 13% higher probability of students meeting their target scores relative to non-users. Relative probabilities were even higher for students starting at the "Unsatisfactory" and "Approaching Basic" levels. Consistent Zearn Math users in these groups were 1.5-1.6x more likely to reach their target score than non-users. For students starting below proficiency, this is an indicator of being on track to proficiency by the 8th grade, meaning consistent Zearn Math users were more likely to be on track to proficiency than non-users (see Results Table 7). Full results are available in Appendix A Table A8.

RESULTS TABLE 7

Probability of Students Meeting their Target Score Across LEAP Achievement Levels

Probabilities of students meeting their spring 2022 LEAP target score for consistent Zearn Math users vs. non-users, by starting Achievement Level

	Consistent Users	Non-Users	Difference	Relative Likelihood
All Students	48.7%	35.8%	12.8%	1.4
Unsatisfactory	66.7%	45.0%	21.6%	1.5
Approaching Basic	62.3%	38.5%	23.8%	1.6
Basic	56.1%	43.1%	13.0%	1.3
Mastery and above	38.2%	29.8%	8.4%	1.3

The relative probabilities of students meeting their target scores between consistent Zearn Math users and non-users were amplified across traditionally disadvantaged subgroups including Black and/or Latino students, economically disadvantaged students, MLLs, students in special education, and chronically absent students. For students in these subgroups, consistent Zearn Math usage was associated with 1.5-1.7x the likelihood of meeting their target scores. Notably, the differences in probability were the most pronounced for students in special education, with consistent Zearn Math users having a 20% higher probability relative to non-users (see Results Table 8). Full results are available in Appendix A Table A9.

RESULTS TABLE 8

Probability of Students Meeting their Target Score Across Subgroups

Probabilities of students meeting their spring 2022 LEAP target score for consistent Zearn Math users vs. non-users, by subgroup

	Consistent Users	Non-Users	Difference	Relative Likelihood
All Students	48.7%	35.8%	12.8%	1.4
Female	47.1%	34.9%	12.3%	1.4
Male	50.1%	36.8%	13.3%	1.4
Black and/or Latino	48.5%	32.1%	16.5%	1.5
Economically disadvantaged	48.6%	31.4%	17.3%	1.6
MLL	51.0%	33.1%	18.0%	1.5
Special education	47.5%	27.7%	19.8%	1.7
Gifted	++	++	++	++
Chronically absent	46.5%	28.7%	17.7%	1.6

++Subgroup does not satisfy baseline equivalence even with statistical adjustment.

Conclusion and Limitations

This analysis provides promising evidence of Zearn Math's positive impact on student achievement. In addition to positive changes in student performance overall, students who started below proficiency, Black and/or Latino students, economically disadvantaged students, MLLs, students in special education, and chronically absent students who consistently used Zearn Math saw even larger gains[19] than the average student. The finding that Zearn Math impacts all students positively, but is associated with even more growth among those starting below proficiency or traditionally disadvantaged students, further substantiates findings from efficacy analyses of Zearn Math's impact in other districts (Zearn 2022a, 2022b, 2022c; Szatrowski, 2022a, 2022b, 2022c; Szatrowski et al., 2022; Rickel, 2023).

By matching students closely on starting LEAP scores in both math and ELA, grade, and 10 demographic and academic factors, treatment and control groups were similar along major confounding characteristics. This technique better isolated the impact of Zearn Math usage as an explanatory factor for differences in academic growth and performance than less rigorous correlational analyses, and meets the WWC criteria for Meets WWC Standards with Reservation and ESSA Tier 2 (see Appendix B for more details). For both students overall and traditionally disadvantaged subgroups, Zearn Math usage appears to drive higher levels of academic growth.

Despite the strong findings from this analysis, there are some limitations. While quasi-experimental methods allow researchers to control for observed confounders, there is a possibility that unobserved confounders mediate the relationship between Zearn Math use and academic performance. Eliminating this limitation entirely would require a randomized controlled trial for Zearn usage.

This study was conducted on a subpopulation of students across 31 parishes in Louisiana. It is possible that the impact of Zearn Math in other locations might show a different effect size, whether larger or smaller. The samples may not be completely representative of Louisiana as a whole. For instance, economically disadvantaged students were underrepresented in the sample relative to the total population across Louisiana, which is a factor that may impact outcomes. It is also possible that there are features specific to these 31 parishes in Louisiana that facilitate large gains with Zearn Math usage that may not be present in other parishes. The geographic specificity of this study may limit its generalizability to a more nationally representative population.

This study's findings of Zearn Math's efficacy align with those from other district and state efficacy analyses (Zearn 2022a, 2022b, 2022c; Szatrowski, 2022a, 2022b, 2022c; Szatrowski et al., 2022; Rickel, 2023). With robust methods and the expansion of efficacy studies to multiple districts across the country, continued replication of trends and findings will provide even stronger evidence of Zearn Math's efficacy moving forward. Zearn plans to continue this work over the coming months and years.

[19] Refers to either scale score gains, percent placing at "Mastery" or above, change in the percent placing at "Mastery" or above, or probability of students meeting their 2022 target scores.

References

Bertsekas, D.P. (1981). A new algorithm for the assignment problem. *Math Programming*, *21*, 152–71. https://doi.org/10.1007/BF01584237

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). cem: Coarsened exact matching in Stata. *Stata Journal*, *9*(4), 524-546. https://doi.org/10.1177/1536867X0900900402

Evan, A., T. Gray, and J. Olchefske.(2006). *The gateway to student success in mathematics and science.* Washington, DC: The American Institutes for Research, with the Microsoft Corporation. https://www.air.org/sites/default/files/downloads/report/Call_for_middle_school_reform_11_1_06_version_0.pdf

Iacus, S. M., King, G., & Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, *106*(493), 345–361. https://doi.org/10.1198/jasa.2011.tm09599

Louisiana Department of Education. (2018). *2018-2019 Grade 4-8 Progress Index Frequently Asked Questions.* LDOE. https://www.louisianabelieves.com/docs/default-source/accountability/2018-2019-grades-4-to-8-progress-index-faq.pdf?sfvrsn=8

Louisiana Department of Education. (2021, August 4). *The Louisiana Department of Education releases new LEAP 2025 data* [Press Release]. LDOE. https://www.louisianabelieves.com/newsroom/news-releases/2021/08/04/the-louisiana-department-of-education-releases-new-leap-2025-data#:~:text=Student%20scores%20are%20reported%20on,for%20the%20next%20grade%20level.

Louisiana Department of Education. (2022a). *2021-2022 Chronic Absenteeism Rate Overall & by Subgroup)* [Data table]. LDOE. https://www.louisianabelieves.com/docs/default-source/data-management/2021-2022-chronic-absenteeism-rate-overall-by-subgroup.xlsx?sfvrsn=d0006218_2

Louisiana Department of Education. (2022b). *2022 Feb SpEd Rates by LEA & Site_PUBLIC* [Data table]. LDOE. https://louisianabelieves.com/docs/default-source/academics/2022-feb-sped-rates-by-lea-site_public.xlsx?sfvrsn=8a486518_6

Louisiana Department of Education. (2022c). *Feb 2022 Multi stats (Total by Site and School System)* [Data table]. LDOE. https://www.louisianabelieves.com/docs/default-source/data-management/feb-2022-multi-stats-(total-by-site-and-school-system).xlsx?sfvrsn=8f546518_2

McCoy, C.E. (2017). Understanding the intention-to-treat principle in randomized controlled trials. *The Western Journal of Emergency Medicine*, *18*(6), 1075–1078. https://doi.org/10.5811/westjem.2017.8.35985

National Center for Education Statistics. (2022). Concentration of public school students eligible for free or reduced-price lunch. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/clb

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education. https://files.eric.ed.gov/fulltext/ED500486.pdf

REL Midwest. (2019). *ESSA tiers of evidence: What you need to know*. Regional Educational Laboratory at American Institutes for Research. https://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/blogs/RELMW-ESSA-Tiers-Video-Handout-508.pdf

Rickel, J. (2023). *Efficacy analysis of Zearn Math: Findings from implementation in a large Southern district.* Zearn. https://webassets.zearn.org/Implementation/LargeSouthernDistrictTechnicalPaper.pdf

Rosenbaum, P. R. (2020). Modern algorithms for matching in observational studies. *Annual Review of Statistics and Its Application*, *7*(1), 143-176. https://dx.doi.org/10.1146/annurev-statistics-031219-041058

Stuart, E. A., & Rubin, D.B. (2008). Matching with multiple control groups with adjustment for group differences. *Journal of Educational and Behavioral Statistics*, *33*(3), 279–306. https://doi.org/10.3102/1076998607306078

Szatrowski, A. (2022a) *Efficacy analysis of Zearn Math: Findings from a large Midwestern urban district.* Zearn. https://webassets.zearn.org/Implementation/MidwesternUrbanDistrictTechnicalPaper.pdf

Szatrowski, A. (2022b) *Efficacy analysis of Zearn Math in Nebraska*. Zearn. https://webassets.zearn.org/Implementation/NebraskaTechnicalAppendix.pdf

Szatrowski, A. (2022c) *Efficacy analysis of Zearn Math in Tulsa*. Zearn. https://webassets.zearn.org/Implementation/TulsaTechnicalPaper.pdf

Szatrowski, A., Rickel, J., & Rosemond, C. (2022) *Efficacy Analysis of Zearn Math in DC Public Schools.* Zearn. https://webassets.zearn.org/Implementation/DCPSTechnicalPaper.pdf

What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook, version 4.1*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5.0-0-508.pdf

Zearn. (2022a). *Consistent Zearn usage dramatically reduces learning loss.* Zearn. https://webassets.zearn.org/Implementation/Zearn_Impact_for_Students_Below_Grade_Level.pdf

Zearn. (2022b). *Economically disadvantaged students, multilingual learners, and Black and Latino students experience double-digit increases in proficiency on 2022 end-of-year assessments with consistent Zearn usage*. Zearn. https://webassets.zearn.org/Implementation/EfficacyResearch_Zearn_Impact_for_Student_Subgroups.pdf

Zearn. (2022c). *For students who scored below grade level in math, consistent Zearn usage tied to 2 grade levels of growth in 2 years of pandemic learning—nearly double the gains of curriculum alone.* Zearn. https://webassets.zearn.org/Implementation/Zearn_Impact_for_Students_Below_Grade_Level.pdf

Appendix A

Table A1

Breakdown of sample matching characteristics		
	Treatment	**Control**
Total *Ns*	7,116	7,116
Pre-scores (Spring '21 assessment scores)		
Math scale score	747.13	746.83
ELA scale score	752.10	752.10
Starting Achievement Level (*Ns*)		
Unsatisfactory (Level 1)	504	522
Approaching Basic (Level 2)	1,084	1,067
Basic (Level 3)	1,894	1,932
Mastery (Level 4)	3,233	3,209
Advanced (Level 5)	401	386
Grade Level (*Ns*)		
Grade 4	4,328	4,328
Grade 5	2,597	2,597
Grade 6	121	121
Grade 7	70	70
Demographic & academic subgroups (*Ns*)		
Female	3,491	3,505
Male	3,625	3,611
Black and/or Latino	2,864	2,881
Economically disadvantaged	2,325	2,550
MLL	145	121
Special education	455	412
Gifted	209	149
Chronically absent	523	491

Table A2

LDOE Schools: comparison of sample and statewide school population			
	Sample-Treatment	**Sample-Control**	**State***
Demographic & academic subgroups			
Female	49%	49%	48%
Male	51%	51%	52%
Black and/or Latino	40%	40%	52%
Economically disadvantaged	33%	36%	71%
MLL	2%	2%	3%
Special education	7%	6%	13%
Gifted	3%	2%	<5%
Chronically absent	7%	7%	19%

*These percentages are based off of the total student population across the 31 parishes represented in this analysis.

Table A3

Spring 2021 and spring 2022 LEAP scale score means, across all students and by subgroup							
	Treatment spring 2021	**Treatment spring 2022**	**Control spring 2021**	**Control spring 2022**	**Starting mean difference**	**Pooled *SD***	**Difference in *SDs****
All Students							
Math scale score	747.13	752.77	746.83	746.20	0.30	29.85	0.01
Starting Proficiency							
Below-proficient ("Unsatisfactory," "Approaching Basic," and "Basic")	722.50	732.82	722.60	724.73	-0.10	19.98	-0.01
Mastery and above ("Mastery" and "Advanced")	770.68	771.88	770.56	767.24	0.12	15.22	0.01
Starting Achievement Level							
Unsatisfactory	686.22	708.70	686.67	698.50	-0.45	10.21	-0.04
Approaching Basic	713.20	725.59	713.24	716.18	-0.03	7.28	0.00
Basic	737.56	743.38	737.49	736.53	0.08	7.11	0.01
Mastery	766.99	769.01	766.95	764.44	0.04	11.20	0.00
Advanced	800.44	795.00	800.59	790.51	-0.15	9.84	-0.01
Grade Level							
Grade 4	746.82	752.57	746.51	746.62	0.31	29.59	0.01
Grade 5	746.71	751.81	746.47	744.76	0.24	30.23	0.01
Grade 6	751.64	760.31	750.97	746.51	0.68	28.69	0.02
Grade 7	774.17	787.83	773.41	773.64	0.76	19.84	0.04

*According to WWC, baseline differences <.05 of a standard deviation satisfy baseline equivalence without adjustment. Differences <.25 of a standard deviation satisfy baseline equivalence with adjustment of difference-in-difference (2022).

Table A3 cont.

Cont. Spring 2021 and spring 2022 LEAP scale score means, across all students and by subgroup							
	Treatment spring 2021	**Treatment spring 2022**	**Control spring 2021**	**Control spring 2022**	**Starting mean difference**	**Pooled *SD***	**Difference in *SDs****
Demographic & academic subgroups							
Female	746.41	751.11	746.48	745.30	-0.07	29.73	0.00
Male	747.82	754.37	747.17	747.08	0.65	29.96	0.02
Black and/or Latino	734.64	741.32	733.07	732.05	1.57	29.20	0.05
Economically disadvantaged	734.00	741.16	733.58	732.09	0.42	30.42	0.01
MLL	723.87	731.81	720.74	720.07	3.13	25.99	0.12
Special education	726.90	733.75	721.18	719.73	5.72	30.19	0.19
Gifted	784.82	792.09	789.37	790.46	-4.55	18.18	.25+
Chronically absent	739.20	744.66	733.49	729.47	5.71	29.51	0.19

*According to WWC, baseline differences <.05 of a standard deviation satisfy baseline equivalence without adjustment. Differences <.25 of a standard deviation satisfy baseline equivalence with adjustment of difference-in-difference (2022).
++Subgroup does not satisfy baseline equivalence even with statistical adjustment.

Table A4

Spring 2021 and spring 2022 adjusted LEAP scale score means, across all students and by subgroup			
	Treatment and control spring 2021 adjusted baseline mean*	**Treatment adjusted spring 2022****	**Control adjusted spring 2022****
All Students			
Math scale score	746.98	752.62	746.35
Starting Proficiency			
Below-proficient	722.55	732.87	724.67
Mastery and above	770.62	771.82	767.30
Starting Achievement Level			
Unsatisfactory	686.45	708.92	698.27
Approaching Basic	713.22	725.60	716.16
Basic	737.53	743.34	736.57
Mastery	766.97	768.99	764.46
Advanced	800.52	795.07	790.43
Grade Level			
Grade 4	746.66	752.41	746.77
Grade 5	746.59	751.69	744.88
Grade 6	751.31	759.97	746.85
Grade 7	773.79	787.45	774.02

*The adjusted baseline mean was calculated as the average of treatment and control students' scale score in spring 2021. See Appendix A Table A3 for the unadjusted 2021 scale scores.
**The adjusted spring 2022 scale scores for treatment and control were calculated by adding the actual score change between spring 2021 to 2022 for each respective group to the spring 2021 adjusted baseline mean. See Appendix A Tables A6 and A7 for the actual change in scale score.

Table A4 cont.

Cont. Spring 2021 and spring 2022 adjusted LEAP scale score means, across all students and by subgroup			
	Treatment and control spring 2021 adjusted baseline mean*	**Treatment adjusted spring 2022****	**Control adjusted spring 2022****
Demographic & academic subgroups			
Female	746.45	751.14	745.26
Male	747.50	754.04	747.40
Black and/or Latino	733.85	740.54	732.84
Economically disadvantaged	733.79	740.95	732.29
MLL	722.31	730.25	721.63
Special education	724.04	730.88	722.59
Gifted	++	++	++
Chronically absent	736.35	741.80	732.32

*The adjusted baseline mean was calculated as the average of treatment and control students' scale score in spring 2021. See Appendix A Table A3 for the unadjusted 2021 and 2022 scale scores.

**The adjusted spring 2022 scale scores for treatment and control were calculated by adding the actual score change between spring 2021 to 2022 for each respective group to the spring 2021 adjusted baseline mean. See Appendix A Tables A6 and A7 for the actual change in scale score.

++Subgroup does not satisfy baseline equivalence even with statistical adjustment.

Table A5

Spring 2021 and spring 2022 percent placing at "Mastery" or above, across all students and by subgroup							
	Treatment spring 2021	**Treatment spring 2022**	**Control spring 2021**	**Control spring 2022**	**Treatment and control spring 2021 adjusted baseline mean***	**Treatment adjusted spring 2022****	**Control adjusted spring 2022****
All Students							
Math percent placing at "Mastery" or above	51.1%	56.0%	50.5%	48.1%	50.79%	55.75%	48.42%
Grade Level							
Grade 4	51.13%	57.05%	50.79%	49.95%	50.96%	56.87%	50.13%
Grade 5	49.75%	52.91%	48.94%	44.24%	49.35%	52.50%	44.65%
Grade 6	52.89%	62.81%	52.07%	43.80%	52.48%	62.40%	44.21%
Grade 7***	92.86%	97.14%	90.00%	88.57%	91.43%	95.71%	90.00%
Demographic & academic subgroups							
Female	49.90%	53.02%	49.24%	46.53%	49.57%	52.69%	46.86%
Male	52.19%	58.92%	51.76%	49.71%	51.98%	58.71%	49.93%
Black and/or Latino	32.82%	38.97%	30.89%	28.67%	31.86%	38.00%	29.64%
Economically disadvantaged	34.15%	38.75%	32.90%	30.08%	33.53%	38.13%	30.70%
MLL***	22.07%	22.07%	13.22%	10.74%	17.65%	17.65%	15.17%
Special education	26.15%	30.11%	17.72%	16.26%	21.94%	25.89%	20.48%
Gifted	++	++	++	++	++	++	++
Chronically absent	40.73%	46.08%	32.38%	26.07%	36.55%	41.91%	30.24%

*The adjusted baseline mean was calculated as the average percent of treatment and control students who placed at "Mastery" or above in spring 2021.
**The adjusted spring 2022 values for treatment and control were calculated by adding the actual percentage point change between spring 2021 to 2022 for each respective group to the spring 2021 adjusted baseline mean.
***The change in percent placing at "Mastery" or above was not significant and therefore not reported in the text. Full results are available in Appendix A Tables A6 and A7.
++Subgroup does not satisfy baseline equivalence even with statistical adjustment.

Table A6

Comparison of changes in scores and percent placing at "Mastery" or above between consistent Zearn Math users and non-users, across all students

	Treatment change in mean	Control change in mean	Difference	Pooled *SD*	Cohen's *d*
All Students					
Math scale score (SS)	5.64	-0.63	6.27***	23.06	0.27
Math percent placing at "Mastery" or above	4.96%	-2.37%	7.34%***	0.51	0.14
Starting Proficiency					
Below-proficient SS	10.27	2.12	8.15***	19.08	0.43
Mastery and above SS	1.20	-3.33	4.52***	16.60	0.27
Starting Achievement Level					
Unsatisfactory SS	22.48	11.82	10.65***	19.50	0.55
Approaching Basic SS	12.38	2.94	9.44***	18.19	0.52
Basic SS	5.82	-0.95	6.77***	18.19	0.37
Mastery SS	2.02	-2.51	4.53***	16.41	0.28
Advanced SS	-5.45	-10.08	4.64***	16.63	0.28
Grade Level					
G4 SS	5.75	0.11	5.64***	18.39	0.31
G4 percent placing at "Mastery" or above	5.91%	-0.83%	6.75%***	0.43	0.16
G5 SS	5.10	-1.71	6.81***	18.04	0.38
G5 percent placing at "Mastery" or above	3.16%	-4.70%	7.86%***	0.42	0.19
G6 SS	8.66	-4.45	13.12***	16.26	0.81
G6 percent placing at "Mastery" or above	9.92%	-8.26%	18.18%***	0.36	0.50
G7 SS	13.66	0.23	13.43***	16.46	0.82
G7 percent placing at "Mastery" or above	4.29%	-1.43%	5.71%	0.21	0.28

*p < .05 **p < .01 ***p < .001

Table A7

Comparison of changes in scores and percent placing at "Mastery" or above between consistent Zearn Math users and non-users, by subgroup

	Treatment change in mean	Control change in mean	Difference	Pooled *SD*	Cohen's *d*
Subgroup					
Female SS	4.70	-1.18	5.88***	18.05	0.33
Female percent placing at "Mastery" or above	3.12%	-2.71%	5.83%***	0.42	0.14
Male SS	6.55	-0.10	6.64***	18.36	0.36
Male percent placing at "Mastery" or above	6.73%	-2.05%	8.78%***	0.42	0.21
Black and/or Latino SS	6.68	-1.01	7.70***	18.86	0.41
Black and/or Latino percent placing at "Mastery" or above	6.15%	-2.22%	8.37%***	0.42	0.20
Economically disadvantaged SS	7.16	-1.50	8.66***	19.43	0.45
Economically disadvantaged percent placing at "Mastery" or above	4.60%	-2.82%	7.43%***	0.41	0.18
MLL SS	7.94	-0.68	8.62***	18.64	0.46
MLL percent placing at "Mastery" or above	0.00%	-2.48%	2.48%	0.37	0.07
Special education SS	6.85	-1.45	8.30***	19.23	0.43
Special education percent placing at "Mastery" or above	3.96%	-1.46%	5.41%*	0.33	0.16
Gifted SS	++	++	++	++	++
Gifted percent placing at "Mastery" or above	++	++	++	++	++
Chronically absent SS	5.46	-4.03	9.48***	19.19	0.49
Chronically absent percent placing at "Mastery" or above	5.35%	-6.31%	11.67%***	0.39	0.30

*p < .05 **p < .01 ***p < .001

++Subgroup does not satisfy WWC standards for baseline equivalence even with statistical adjustment.

Table A8

Logistic regression probabilities of students meeting their 2022 target scores between consistent Zearn Math users and non-users, across all students

	Treatment mean	Control mean	Difference	Relative likelihood	Odds ratio
All Students					
Probability of meeting target score	48.65%	35.83%	12.82%	1.36	1.69***
Starting Proficiency					
Below-proficient probability of meeting target	59.53%	42.01%	17.53%	1.42	2.03***
Mastery and above probability of meeting target	38.22%	29.79%	8.43%	1.28	1.46***
Starting Achievement Level					
Unsatisfactory probability of meeting target	66.67%	45.02%	21.65%	1.48	2.44***
Approaching Basic probability of meeting target	62.27%	38.52%	23.75%	1.62	2.63***
Basic probability of meeting target	56.07%	43.12%	12.96%	1.30	1.68***
Mastery probability of meeting target	36.53%	28.17%	8.36%	1.30	1.47***
Advanced probability of meeting target	51.87%	43.26%	8.61%	1.20	1.41*
Grade Level					
G4 probability of meeting target	49.61%	38.45%	11.16%	1.29	1.58***
G5 probability of meeting target	46.86%	32.15%	14.71%	1.46	1.86***
G6 probability of meeting target	49.59%	22.31%	27.27%	2.22	3.42***
G7 probability of meeting target	54.29%	34.29%	20.00%	1.58	2.28***

*p < .05 **p < .01 ***p < .001

Table A9

Logistic regression probabilities of students meeting their 2022 target scores between consistent Zearn Math users and non-users, by subgroup

	Treatment mean	Control mean	Difference	Relative likelihood	Odds ratio
Subgroup					
Female probability of meeting target	47.12%	34.86%	12.26%	1.35	1.66***
Male probability of meeting target	50.12%	36.78%	13.35%	1.36	1.73***
Black and/or Latino probability of meeting target	48.53%	32.07%	16.46%	1.51	2.00***
Economically disadvantaged probability of meeting target	48.65%	31.37%	17.27%	1.55	2.07***
MLL probability of meeting target	51.03%	33.06%	17.98%	1.54	2.11**
Special education probability of meeting target	47.47%	27.67%	19.80%	1.72	2.36***
Gifted probability of meeting target	++	++	++	++	++
Chronically absent probability of meeting target	46.46%	28.72%	17.75%	1.62	2.15***

*p < .05 **p < .01 ***p < .001

++Subgroup does not satisfy WWC standards for baseline equivalence even with statistical adjustment.

Appendix B

This study was designed to meet the What Works Clearinghouse (WWC) “Meets WWC Group Design Standards with Reservations'' rating and to meet an Every Student Succeeds Act (ESSA) Tier 2 (Moderate) rating on the ESSA guidelines for evidence-based interventions. This Appendix provides more detail about the criteria for these designations and how this impact study meets those criteria.

What Works Clearinghouse provides ratings of randomized control trials (RCTs) and quasi-experimental designs (QEDs) against their Group Design standards. There are three possible ratings: Meets WWC Standards without Reservations, Meets WWC Standards with Reservations, or Does Not Meet WWC Standards. Because QED studies that establish baseline equivalence or use acceptable statistical adjustments “reduce, but likely do not eliminate, the potential bias associated with the group assignment procedures,” Meets WWC Standards with Reservations is the highest possible rating for QEDs (What Works Clearinghouse, 2022).

This study uses quasi-experimental matching methods to create baseline equivalency between treatment and control groups along major confounding factors. Consistent Zearn Math users were matched with non-users, in the same grade, on starting math and English Language Arts (ELA) standardized test scores, along with ten student characteristics, using a two-step Coarsened Exact Matching (CEM) method with optimal matching. CEM is a technique that simulates block sampling by matching students on covariates related both to a student’s likelihood of using Zearn Math consistently and their academic performance (Blackwell et al., 2009; Iacus et al., 2011).

A QED study must satisfy several criteria to meet the WWC standard of “Meets WWC Standards with Reservations.” The first is that the outcome measure “meets four standards: (1) face validity, (2) reliability, (3) not over aligned with the intervention, and (4) consistent data collection procedures” (What Works Clearinghouse, 2022). In this study, the primary outcome is math achievement on the Louisiana Education Assessment Program (LEAP). WWC considers standardized tests that are routinely administered in educational settings, like the LEAP, to meet these standards (2022).

The next criteria is the elimination of confounding factors (What Works Clearinghouse, 2022). By matching fidelity users to non-users within five scale score points on their pre-score for both math and ELA on the LEAP, as well as at least six of ten other student characteristics — parish, school, use case, gender, race, poverty level, multilingual learner status (MLL), special education status, gifted status, and chronic absenteeism status — the design of this study creates two groups that are academically and demographically similar on the most relevant and measurable confounding factors that would impact academic growth.

While CEM allows researchers to control for observed confounders, a possibility exists that there are unmeasured factors that differentiate the comparison groups of students who reach fidelity and those with no usage. For example, it is possible that an unmeasured characteristic allows fidelity users to reach higher usage than would be possible for non-users. However, this type of unmeasurable

attribute is what WWC refers to as “imperfect overlap in the characteristic between the conditions,” which they term a selection mechanism, not a confounding factor (2020, p. 82).

This possibility of an unmeasured characteristic that could bias estimates is similar to an example provided by WWC of a program based on voluntary enrollment in which students who volunteer could differ from those who did not in hard-to-measure qualities like introversion vs. extroversion. It clarifies that “the WWC does not consider this to be a confounding factor, but the selection mechanism and potential difference in unmeasured characteristics are reasons that QEDs are limited to a rating of Meets WWC Group Design Standards with Reservations, if the baseline equivalence requirement is satisfied” (2020, p. 82).

The final criteria for a quasi-experimental study to meet WWC Standards with Reservations is illustrating baseline equivalence between treatment and control groups. This can be done with a pre-intervention measure that is the same as the outcome measure (2022). In this case, LEAP math scores are used as a pre-intervention measure of baseline equivalence and as the outcome measure of the study.

According to WWC, baseline differences < .05 of a standard deviation satisfy baseline equivalence without adjustment. Differences < .25 of a standard deviation satisfy baseline equivalent with statistical adjustment. Difference-in-difference is an acceptable statistical adjustment (2022). All groups in this study meet the criteria for baseline equivalence either without or with adjustment, with the exception of gifted students (see Appendix B Table B1). Results for this subgroup are not reported in the results as they do not qualify as baseline equivalent even with statistical adjustment.

Table B1_

Baseline Equivalence by Subgroup

Study qualifications for WWC baseline equivalence standards*	
All students	Meets
Grades	Meets
Achievement Levels	Meets
Female	Meets
Male	Meets
Black and/or Latino	Meets w/adjustment
Economically disadvantaged	Meets
MLL	Meets w/adjustment
Special education	Meets w/adjustment
Gifted	Does not meet
Chronically absent	Meets w/adjustment

*Baseline differences < .05 of a standard deviation satisfy baseline equivalence without adjustment. Differences < .25 of a standard deviation satisfy baseline equivalent with statistical adjustment. For baseline equivalence data, see Appendix A Table A3.

WWC Essa Tier 2 designation requires a strong quasi-experimental research design that would qualify for Meets WWC Standards with Reservations. In addition, an ESSA Tier 2 rating requires a minimum of 350 students. This analysis has a sample size of 7,116 students in each group for a total of 14,232, which exceeds 350. In addition, the study must have been conducted in more than one school. This study spans 228 treatment schools, with an additional 223 of the 446 total control schools.

Finally, findings must be statistically significant and there can be “no strong negative findings from experimental or quasi-experimental studies” (Regional Educational Laboratory at American Institutes for Research, 2019, p. 2). Results from this study show statistically significant positive impacts from the implementation of Zearn Math. There have been no strong negative findings from other experimental or quasi-experimental studies, while there have been statistically significant positive findings from other QED Zearn studies (see 2022a, 2022b, 2022c; Szatrowski, 2022a, 2022b, 2022c; Szatrowski et al., 2022; Rickel, 2023).